The FHWA Travel Model Improvement Program Workshop over the Web

The Travel Model Development Series: Part I – Travel Model Estimation

presented by
Thomas Rossi
Yasasvi Popuri
Cambridge Systematics, Inc.

December 11, 2008

1

Webinar Outline

- Session 1: Introduction – October 16, 2008
- Session 2: Data Set Preparation – November 6, 2008
- Session 3: Estimation of Non-Logit Models – December 11, 2008
- Session 4: Estimation of Logit Models – February 10, 2009

2

Webinar Outline (continued)

- Session 5: Application and Validation of Logit Models – March 12, 2009
- Session 6: Advanced Topics in Discrete Choice Models – April 14, 2009
- Session 7: Trip Assignment – May 7, 2009
- Session 8: Evaluation of Validation Results – June 9, 2009

3

Homework
From Session 2

4

Aggregate Model Components
Four-Step Models

- **Trip production**
- **Trip attraction**
- **Trip distribution**
- Mode choice
- Assignment
- **Time of day**
- Auto ownership
- Other
  - Trucks/freight
  - External trips
  - Other?

5

Why Is Getting the Parameters Called Model "Estimation"?

- We cannot know the true parameters, and so we must estimate their values
- Every person has his/her own parameters
- We use the same parameters for groups of similar travelers

6

Relevant Statistical Concepts

- Estimators
- Maximum likelihood
- Confidence intervals
- Statistical significance

7

Choosing the Independent Variables

- Relevance to the travel choice
- Availability in the estimation data set
- Availability for model application (forecasting)
- Statistical testing

8

Trip Data File
From Household/On-Board Surveys

- Each record represents a trip made by an individual
- Each field represents a characteristic of:
  - The trip;
  - The traveler;
  - His/her household; or
  - The areas traveled

9

Trip Data File
Typical Fields

From the survey

- Origin and destination
- Trip purpose
- Chosen mode
- Time of day of trip
- *Trip time/cost*
- Household/person characteristics (linked from household/person file)

From other sources

- Travel time (in-vehicle)
- Other time components (wait, access/egress, transfer)
- Costs (parking, auto operating, transit fare)
- Number of transit transfers
- Zone attributes
- Logsums from other models

10

Should Weighted Data Be Used in Estimating Aggregate Models?

- Example:

Household	Trips	Sampled?	Weight
1	5	Yes	1.0
2	10	Yes	2.0
3	10	No	(not surveyed)

11

Typical Aggregate Model Types

- **Trip production**
  - **Cross-classification**
- **Trip attraction**
  - **Linear regression**
- **Trip distribution**
  - **Gravity**
- **Time of day**
  - **Simple factoring**

12

Cross-Classification Model

Independent Variable #2	Independent Variable #1				
		Value 1	Value 2	…	Value n
	Value 1	Dep var value	Dep var value		Dep var value
	Value 2	Dep var value	Dep var value		Dep var value
	…				
	Value n	Dep var value	Dep var value		Dep var value

13

Practical Considerations in Trip Production Model Estimation

- At most two, possibly three variables can be used
- Different cross-classifications by purpose?
- Maximum likelihood estimator: Mean trip rate per household

14

Estimating the Cross-Classification Trip Production Model

1. Decide on variables

15

Trip Production Model
Typical Variables (Trip Purpose)

- **Number of persons (any)**
- **Number of workers (work, possibly others)**
- **Number of children (school)**
- **Number of autos (any)**
- **Income level (any)**

16

Estimating the Cross-Classification Trip Production Model

1. Decide on variables
2. Compute average trip rates by cross-classified variables
   - Using database manager, statistical software, or spreadsheet
3. Check cells for logical relationships
4. Test alternative combinations of variables

17

Example
Home Based Work Productions

Workers	Autos				
		0	1	2	3+
	0	0.00	0.00	0.00	2.00
	1	1.50	1.51	1.54	1.54
	2	2.98	2.98	3.02	3.04
	3+	5.09	4.29	4.91	4.91

18

Example
Home Based Work Productions

Workers	Autos				
		0	1	2	3+
	0	0.00	0.00	0.00	2.00
	1	1.50	1.51	1.54	1.54
	2	2.98	2.98	3.02	3.04
	3+	5.09	4.29	4.91	4.91

19

Example
Home Based Work Productions

Workers	Autos				
		0	1	2	3+
	0	0.00	0.00	0.00	0.00
	1	1.50	1.51	1.54	1.54
	2	2.98	2.98	3.02	3.04
	3+	4.69	4.69	4.91	4.91

20

Regression Model

$Y = B_0 + B_1 X_1 + B_2 X_2 + \ldots + B_n X_n$

where:

Y = Dependent variable

B_i = Estimated coefficients

X_i = Independent variables

The maximum likelihood estimators for coefficients are based on method of least squares

21

Setting up Data for Trip Attraction Model Estimation
Linear Regression

- Define independent variables to be tested
- Use trip file – weighted data
- Define dependent variable (i.e. number of trips for the purpose being analyzed)
- Aggregate to districts
- Attach district level data
  - Employment by type
  - Households

22

Example Data Set
Trip Attraction Model Estimation

ID	Purpose	NUMTRIPS	ATTR	ADIST	ADISTNAME	WEIGHT
11929	1	1	107	1	CBD	107.1746
11929	1	1	107	1	CBD	107.1746
11929	1	1	107	1	CBD	107.1746
11929	1	1	107	1	CBD	107.1746
11982	1	1	114	1	CBD	99.65
11982	1	1	114	1	CBD	99.65
12747	1	4	113	1	CBD	99.65

23

Example Data Set
Trip Attraction Model Estimation

Attraction District	Total JTW Attractions	Total Employment
Airport	62,753	48,839
CBD	80,800	58,085
Tonyville	10,929	15,290
Christown	19,041	20,728
Samville	56,748	40,464
Thomas County	626	1,966

24

Estimating the Trip Attraction Model Using Linear Regression

1. Run statistical software to estimate coefficients
2. Evaluate results
3. Revise specification and reestimate
   - Consider alternative variable definitions, combinations
   - Eliminate variables as appropriate
4. Choose “best” specification

25

Evaluating the Trip Attraction Model Using Linear Regression

1. Reasonableness of coefficient estimates
   - Sign (positive)
   - Magnitudes – marginal contributions and relative values
2. Significance of estimates (t-statistics)
   - $|t| > 1.96$ implies significance at 95% level
3. Goodness of fit (R^2)
   - Range: 0 – 1
   - Typically > 0.9 for most trip purposes

26

Trip Attraction Model
Example Results

HBW Attr = 560.0 (0.3) + 1.20 (20.0)*Employment

(t-statistics in parentheses)

$R^2 = 0.946$

27

Gravity Model
Trip Distribution

$$T_{ij} = \frac{P_i A_j F(t)_{ij} K_{ij}}{\sum_j P_i A_j F(t)_{ij} K_{ij}}$$

where:

- T_{ij} = number of trips produced in zone i and attracted to zone j
- P_i = trips produced in zone i
- A_j = trips attracted to zone j
- $F(t)_{ij}$ = friction factor from i to j (based on impedance t)
- K_{ij} = K factor from i to j
- i = origin zone
- j = destination zone

28

Setting up Data for Aggregate Model Estimation
Gravity Model

- Define independent variable (e.g. highway travel time)
- Use trip file – weighted data
- Compute trip length frequency distribution by trip purpose

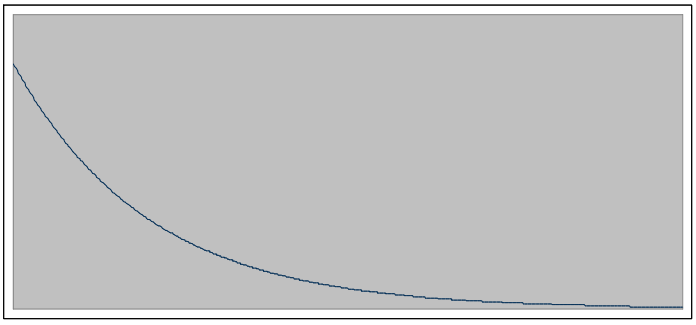
29

Model Estimation
Gravity Model

- Typical methods:
  - Use friction factor fitting in modeling software
  - Simple function with known maximum likelihood parameter
    Example: Exponential distribution with parameter 1/M where M = sample mean travel time
  - Function with parameters transferred from other model
    Example: Gamma distribution

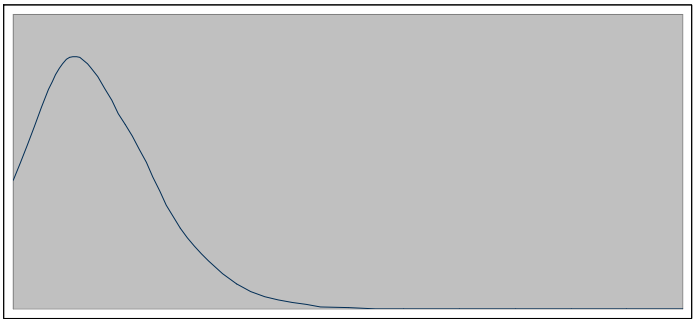
30

Exponential Distribution



31

Gamma Distribution



32

Developing Time of Day Factors

- What is desired:
  - Percentages of daily trips for each purpose that occur in each period, by direction for home based trips (non-directional for non-home based trips)

33

Setting up Data for Aggregate Model Estimation
Time of Day Model

- Determine resolution for testing (e.g. half hours)
- Use trip file – weighted data
- Define time variable (e.g. departure time, arrival time, midpoint)
- Define time periods

34

Defining the Peak Periods
Example

Peaks	Hours	% of Daily Trips	Add'l %
0.5 hr	8:00-8:30	4.1%	4.1%
1 hr	7:30-8:30	7.9%	3.9%
1.5 hr	7:30-9:00	11.5%	3.5%
2 hr	7:00-9:00	14.6%	3.1%
2.5 hr	7:00-9:30	16.9%	2.3%
3 hr	7:00-10:00	19.0%	2.1%

35

Example Time of Day Factors

		AM Peak Period (7:00 – 9:00)	PM Peak Period (3:00 – 6:00)
HBW	From home	26.3%	4.1%
	To home	1.1%	21.3%
HBNW	From home	10.7%	12.3%
	To home	1.5%	10.7%
NHB		4.3%	22.2%

36

Homework
Session 3

37